TITLE

MUTATIONS AFFECTING PLASMID COPY NUMBER

This application claims the benefit of U.S. Provisional Application No. 60/434,973 filed December 20, 2002.

FIELD OF THE INVENTION

This invention is in the field of microbiology. More specifically, this invention pertains regulating copy number of pBR and pACYC based plasmids.

BACKGROUND OF THE INVENTION

Molecular biotechnology is a discipline that is based on the ability of researchers to transfer specific units of genetic information from one organism to another. This process, known as cloning, relies on the techniques of recombinant DNA technology to produce a useful product or a commercial process (Glick, B. R.; Pasternak, J. J., Molecular Biotechnology Principles and Applications of Recombinant DNA, 2nd ed. American Society for Microbiology, Washington, DC. 1998).

Commercial processes often require that proteins encoded by the cloned genes are produced at high rates of expression. There is no single strategy for achieving maximal expression of every cloned gene. Most cloned genes have distinctive molecular properties that require the investment of considerable time and effort before a specific set of conditions that result in an appropriate level of expression is found. There are a variety of ways to modulate gene expression. Microbial metabolic engineering generally involves the use of multi-copy vectors to express a gene of interest under the control of a strong or conditional promoter. Increasing the copy number of cloned genes generally increases amounts and activity of encoded enzymes, therefore allowing increased levels of product formation that is important to commercial processes. However, it is sometimes difficult to maintain vectors in host cells due to instability. Deleterious effects on cell viability and growth can be observed due to the vector burden. The introduction and expression of foreign DNA in a host organism often changes the metabolism of the organism in ways that may impair normal cellular functioning. This phenomenon is due to a metabolic load or burden imposed upon the host by the foreign DNA. The metabolic load may result from a variety of conditions including: 1) increasing plasmid copy number, 2) overproduction of proteins, 3) saturation of export sites, and/or 4) interference of cellular function by the foreign protein itself. It is also

difficult to control the optimal expression level of desired genes on a vector. Several reports have suggested altering the copy number of plasmids can have benefit in production of recombinant protein and analysis of transcriptional fusions (Grabherr et al., *Biotech. Bioeng.*, 77:142-147 (2002); Podkovyrov, S. M. and Larson, T. J., *Gene,* 156:151-152 (1995)).

Bacterial plasmids are extrachromosomal genomes that replicate autonomously and in a controlled manner. Many plasmids are self-transmissible or mobilizable by other replicons, thus having the ability to colonize new bacterial species. In nature, plasmids may provide the host with valuable functions, such as drug resistance(s) or metabolic pathways useful under certain environmental conditions, although they are likely to constitute a slight metabolic burden to the host. To co-exist stably with their hosts and minimize the metabolic load, plasmids must control their replication, so that the copy number of a given plasmid is usually fixed within a given host and under defined cell growth conditions.

The number of copies of a plasmid can vary from 1, as in the case of the F plasmid, to over a hundred for pUC18. Bacterial plasmids maintain their number of copies by negative regulatory systems that adjust the rate of replication per plasmid copy in response to fluctuations in the copy number. Three general classes of regulatory mechanisms have been studied in depth, namely those that involve directly repeated sequences (iterons), those that use only antisense RNAs (AS-RNA), and those that use a mechanism involving an antisense RNA in combination with a protein.

Several chromosomal genes are known to affect the copy number of certain groups of plasmids. The *pcnB* gene encoding the poly(A) polymerase I has been found to affect copy number of ColE1 plasmids in *Escherichia coli*. Mutations in the *pcnB* locus of *E. coli* reduce the copy number of ColE1-like plasmids, which include pBR322-derived plasmids (Lopilato et al., *Mol. Gen. Genet.*, 205:285-290 (1986)) and pACYC-derived plasmids (Liu et al., *J. Bacteriol.*, 171:1254-1261 (1989)). Furthermore, it was discovered that the *pcnB* gene product was required for copy number maintenance of ColE1 and R1 plasmids of the IncFII compatibility group. Copy number of R1 plasmids like ColE1 is controlled by an antisense RNA mechanism, though the mechanism is different between the two. The iteron-regulated plasmids F and P1 were maintained normally in strains deleted for *pcnB*.

The gene *relA* encoding (p)ppGpp synthetase 1 allows cells to initiate stringent response during starvation. ColE1-type of plasmids can be amplified in amino acid-starved *relA* mutants of *Escherichia coli* (Wrobel et al., *Microbiol Res.*, 152:251-255 (1997)). Differential amplification efficiency of plasmids pBR328 (pMB1-derived replicon) and pACYC184 (p15A-derived replicon) was observed in the *relA* mutant during starvation for particular amino acids.

A recent paper described an origin-specific reduction of ColE1 plasmid copy number due to specific mutations in a distinct region of *rpoC* (Ederth et al., *Mol. Gen. Genomics*, 267:587-592 (2002)). The specific mutations, including a single amino acid substitution (G1161R) or a 41-amino acid deletion (Δ1149-1190), are located near the 3'-terminal region in the *rpoC* gene, encoding the largest subunit β' of the RNA polymerase. These mutations cause over 20- and 10-fold reductions, respectively, in the copy number of ColE1. The RNA I/RNA II ratio, which controls the ColE1 plasmid copy number, was affected by these mutations.

The problem to be solved is to identify and provide chromosomal gene modifications that alter plasmid copy number in bacteria. The present invention has solved the stated problem through the discovery that disruptions in any one of 5 (*thrS, rpsA, rpoC, yjeR,* and *rhoL*) chromosomal genes will result in increase of copy number of certain plasmids. The effect of mutation of these loci on plasmids is novel and could not have been predicted from known studies.

SUMMARY OF THE INVENTION

The invention provides bacterial production host comprising:

a) a plasmid comprising:

(i) a target gene to be expressed; and

(ii) a replicon controlled by antisense-RNA regulation; and

b) a mutation in a gene selected from the group consisting of *thrS, rpsA, rpoC, yjeR, and rhoL* wherein the nucleotide sequence of the mutated *thrS* gene is SEQ ID NO:19; the nucleotide sequence of the mutated *rpsA* gene is SEQ ID NO:21; the nucleotide sequence of the mutated *rpoC* gene is SEQ ID NO:22; the nucleotide sequence of the mutated *yjeR* gene is SEQ ID NO:23; and the sequence of the mutated *rhoL* gene is SEQ ID NO:25.

In a preferred embodiment the invention provides a method for the expression of a target gene comprising:

a) providing an bacterial production host of the invention comprising a target gene to be expressed;

b) growing the production host of step (a) under suitable conditions wherein the target gene is expressed.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

Figure 1 shows the strategy for mutagenesis and screening of *E. coli* chromosomal mutants that affect carotenoid production.

Figure 2 shows the β-carotene production in *E. coli* mutants.

Figure 3 is an image of a gel electrophoresis showing the amount of plasmid DNA isolated from the carotenoid-synthesizing plasmid pPCB15 isolated from wild type MG1655 and the mutants that affected carotenoid production.

Figure 4 an image of a gel electrophoresis showing levels of plasmid DNA extracted from mutants showing increased carotenoid production.

Figure 5 shows the luciferase activity from the *luxCDABE* reporter plasmid pTV200 in MG1655 and the mutants.

Figure 6 is a gel comparing the isolated plasmid DNA of pTV200 as compared with wild type MG1655 and related mutants.

Figure 7 is a gel comparing the isolated plasmid DNA of pBR328 with that from wild type MG1655 and related mutants.

Figure 8 is a gel showing plasmids DNA from different replicons in MG1655 and the W4 and Y15 mutants.

The invention can be more fully understood from the following detailed description and the accompanying sequence descriptions, which form a part of this application.

The following sequences comply with 37 C.F.R. 1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures - the Sequence Rules") and are consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (1998) and the sequence listing requirements of the EPO and PCT (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

Table 1. Nucleotide and amino acid sequences for *Pantoea stewartii* carotenoid biosynthesis genes.

Gene/Protein Product	Source	Nucleotide SEQ ID NO	Amino Acid SEQ ID NO
CrtE	*Pantoea stewartii*	1	2
CrtX	*Pantoea stewartii*	3	4
CrtY	*Pantoea stewartii*	5	6
CrtI	*Pantoea stewartii*	7	8
CrtB	*Pantoea stewartii*	9	10
CrtZ	*Pantoea stewartii*	11	12

SEQ ID NOs:13-14 are oligonucleotide primers used to amplify the carotenoid biosynthetic genes from *P. stewartii*.

SEQ ID NOs:15-18 are oligonucleotide primers used to screen for the Tn5 insertion site in mutants of the present invention.

SEQ ID NO: 19 is the nucleotide sequence of the mutated *thrS* gene with the Tn5 insertion.

SEQ ID NO: 20 is the nucleotide sequence of the mutated *deaD* gene with the Tn5 insertion.

SEQ ID NO: 21 is the nucleotide sequence of the mutated *rpsA* gene with the Tn5 insertion.

SEQ ID NO: 22 is the nucleotide sequence of the mutated *rpoC* gene with the Tn5 insertion.

SEQ ID NO: 23 is the nucleotide sequence of the mutated *yjeR* gene with the Tn5 insertion.

SEQ ID NO: 24 is the nucleotide sequence of the mutated *mreC* gene with the Tn5 insertion.

SEQ ID NO: 25 is the nucleotide sequence of the mutated *rhoL* gene with the Tn5 insertion.

SEQ ID NO: 26 is the nucleotide sequence of the mutated *hscB* (*yfhE*) gene with the Tn5 insertion.

SEQ ID NO: 27 is the nucleotide sequence of the mutated *pcnB* gene with the Tn5 insertion.

SEQ ID NO: 28 is the nucleotide sequence for the plasmid pPCB15.

DETAILED DESCRIPTION OF THE INVENTION

The invention relates to a method for regulating plasmid copy number for plasmids exhibiting anti-sense RNA copy-number control including those under the control of the pMB1 and p15A replicons. Specifically, it has been discovered that mutations in the chromosomal genes *thrS, rpsA, rpoC, yjeR,* and *rhoL* have an effect on plasmid copy number of these plasmids.

The ability to regulate the copy number of plasmids has implications for the production of many microbially produced industrial chemicals and pharmaceuticals where additional copies of key pathway genes will enhance pathway performance.

In this disclosure, a number of terms and abbreviations are used. The following definitions are provided.

"Open reading frame" is abbreviated ORF.

"Polymerase chain reaction" is abbreviated PCR.

The term "p15A" refers to a replicon for a family of plasmid vectors including pACYC-based vectors.

The term "pMB1"refers to a replicon for a family of plasmid vectors including pUC and pBR based vectors

The term "replicon" refers to a genetic element that behaves as an autonomous unit during replication. It contains sequences controlling replication of a plasmid including its origin of replication.

The term "ColE1" refers to a replicon for a family of plasmid vectors including p15A and pMB1.

The term "pACYC derived plasmids" refers to a family of plasmids derived from the p15A origin.

The term "(p)ppGpp synthetase 1" refers to the enzyme coded for by the *relA* gene. (p)ppGpp refers to both guanosine tetraphosphate (ppGpp) and guanosine pentaphosphate (p)ppGpp, unusual nucleotides involved in the stringent response.

The term "stringent response" refers to the cellular response to lack of amino acids necessary for protein synthesis.

The term "iterons" refers to directly repeating DNA sequences located either within or slightly outside of the origin of replication of a plasmid to which regulatory proteins bind to in order to initiate and regulate replication.

The term "RNA I" refers to a 108 nucleotide molecule of RNA, complementary to the 5' end of RNA II, that is a negative regulator of replication of many plasmid origins.

The term "RNA II" refers to an RNA transcript made by RNA polymerase that allows for the initiation of replication of a plasmid.

The terms "anti-sense RNA copy-control" and "AS-RNA" refer to one of the methods by which plasmid copy number is controlled.

The term "production host" means a bacteria engineered to produce a specific genetic end product. The term "enteric production host" means an enteric bacteria engineered to produce a specific genetic end product. Typical examples of enteric bacteria are the genera *Escherichia* and *Salmonella*.

The term "isoprenoid" or "terpenoid" refers to the compounds and any molecules derived from the isoprenoid pathway including 10 carbon terpenoids and their derivatives, such as carotenoids and xanthophylls.

The "Isoprenoid Pathway" as used herein refers to the enzymatic pathway that is responsible for the production of isoprenoids. At a minimum, the isoprenoid pathway contains the genes *dxs, dxr, ygpP(ispD), ychB(ispE), ygbB(ispF), lytB, idi, ispA,* and *ispB* which may also be referred to herein as the "Upper Isoprenoid Pathway". The "Carotenoid Biosynthetic Pathway", "Lower Isoprenoid Pathway" or "Lower Pathway" refers to the genes encoding enzymes necessary for the production of carotenoid compounds and include, but are not limited to *crtE, crtB, crtI, crtY, crtX,* and *crtZ*.

The term "carotenoid biosynthetic enzyme" is an inclusive term referring to any and all of the enzymes encoded by the *Pantoea crtEXYIB* cluster. The enzymes include CrtE, CrtY, CrtI, CrtB, and CrtX.

The term "pPCB15" refers to the pACYC-derived plasmid containing β-carotene synthesis genes *Pantoea crtEXYIB*, used as a reporter plasmid for monitoring β-carotene production in *E. coli*.

The term "*E. coli*" refers to *Escherichia coli* strain K-12 derivatives, such as MG1655 (ATCC 47076) and MC1061 (ATCC 53338).

The term "*Pantoea stewartii*" used interchangeably with *Erwinia stewartii* (Mergaert et al., *Int J. Syst. Bacteriol.*, 43:162-173 (1993)).

The term "*Pantoea ananatas*" is used interchangeably with *Erwinia uredovora* (Mergaert et al., *supra*).

The term "*Pantoea crtEXYIB* cluster" refers to a gene cluster containing carotenoid synthesis genes *crtEXYIB* amplified from *Pantoea stewartii* ATCC 8199. The gene cluster contains the genes *crtE*, *crtX*, *crtY*, *crtI*, and *crtB*. The cluster also contains a *crtZ* gene organized in opposite direction adjacent to *crtB* gene.

The term "CrtE" refers to the geranylgeranyl pyrophosphate synthase enzyme encoded by *crtE* gene which converts trans-trans-farnesyl diphosphate + isopentenyl diphosphate to pyrophosphate + geranylgeranyl diphosphate.

The term "CrtY" refers to the lycopene cyclase enzyme encoded by *crtY* gene which converts lycopene to β-carotene.

The term "CrtI" refers to the phytoene dehydrogenase enzyme encoded by *crtI* gene which converts phytoene into lycopene via the intermediaries of phytofluene, zeta-carotene, and neurosporene by the introduction of 4 double bonds.

The term "CrtB" refers to the phytoene synthase enzyme encoded by *crtB* gene which catalyzes reaction from prephytoene diphosphate (geranylgeranyl pyrophosphate) to phytoene.

The term "CrtX" refers to the zeaxanthin glucosyl transferase enzyme encoded by *crtX* gene which converts zeaxanthin to zeaxanthin-β-diglucoside.

The term "CrtZ" refers to the β-carotene hydroxylase enzyme encoded by the *crtZ* gene which catalyses hydroxylation reaction from β-carotene to zeaxanthin.

The term "pTV200" refers to the plasmid based upon the pACYC184 plasmid that contains a promoterless *luxCDABE* gene cassette from *Photorabdus luminescens* and produces luminescence or light when transformed into *E. coli*.

The term "pBR328" refers to one of the pBR plasmids derived from the pMB1 replicon.

The term "pACAY184" refers to one of the pACYC plasmids derived from the p15A replicon.

The term "pSC101" refers to the representative plasmid belonging to the pSC101 replicon group.

The term "pBHR1" refers to the plasmid derived from the pBBR1 replicon with a broad host range origin of replication.

The term "pMMB66" refers to the plasmid derived from RSF1010 that belongs to the IncQ incompatibility group.

The term "pTJS75" refers to the plasmid derived from RK2 that belongs to the IncP incompatibility group.

The term "*pcnB*" refers to the poly(A) polymerase gene locus.

The term "*thrS*" refers to the threonyl-tRNA synthetase gene locus.

The term "*deaD*" refers to the RNA helicase gene locus.

The term "*rpsA*" refers to the 30S ribosomal subunit protein S1 gene locus.

The term "*rpoC*" refers to the RNA polymerase β' subunit gene locus.

The term "*yjeR*" refers to the oligoribonuclease gene locus.

The term "*mreC*" refers to the rod-shape determining protein gene locus.

The term "*rhoL*" refers to the rho operon leader peptide gene locus.

The terms "*yfhE*" or "*hscB*" refer to the heat-shock-cognate-protein gene locus.

The term "incompatibility group" refers to plasmids that cannot co-exist in a bacterial host. Generally, plasmids within the same incompatibility group have similar mechanisms of replication and replication control.

The term "Rep" refers to the replication proteins that initiate plasmid replication. Many Rep proteins also regulate the frequency of initiation.

The term "Rop" refers to a small protein which when it binds to both RNA molecules, increases the stability of the RNA I/ RNA II complex, thus decreasing the likelihood of plasmid replication.

The term "CopG" refers to a transcriptional repressor protein of plasmid replication.

The term "RNAP" refers to RNA polymerase.

As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

The term "genetic end product" means the substance, chemical or material that is produced as the result of the activity of a gene product. Typically a gene product is an enzyme and a genetic end product is the product of that enzymatic activity on a specific substrate. A genetic end product may the result of a single enzyme activity or the result of a number of linked activities, such as found in a biosynthetic pathway (several enzyme activites).

The term "complementary" is used to describe the relationship between nucleotide bases that are capable to hybridizing to one another. For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine.

"Codon degeneracy" refers to the nature in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments that are then enzymatically assembled to construct the entire gene. "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well-established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" or "exogenous" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric

genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure. "Disrupted gene" refers to a gene fragment disrupted by an insertion of a foreign DNA such as a transposon. Disruption in the 5' end or the middle of the gene likely abolishes the function of the gene. Disruption close to the 3' terminal end of the gene might result in altered function from the truncated protein. "Target gene" is the gene of interest that is used in the synthesis of a desired genetic end product, usually resulting in a measurable phenotypic change in the microorganism.

"Operon", in bacterial DNA, is a cluster of contiguous genes transcribed from one promoter that gives rise to a polycistronic mRNA.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing site, effector binding site and stem-loop structure.

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions ("inducible promoters"). Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". Promoters can be further classified by the relative strength of expression observed by their use (i.e. weak, moderate, or strong). It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence capable of affecting mRNA processing or gene expression.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from post-transcriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (US 5,107,065; WO 99/28508). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that is not translated yet has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic", "recombinant" or "transformed" organisms.

The terms "plasmid", "vector" and "cassette" refer to an extrachromosomal element often carrying genes which are not part of the

central metabolism of the cell, and usually in the form of circular double-stranded DNA fragments. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Transformation cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitate transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for enhanced expression of that gene in a foreign host.

The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. "Sequence analysis software" may be commercially available or independently developed. Typical sequence analysis software will include but is not limited to the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990), and DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 USA), and the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.,* [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Publisher: Plenum, New York, NY. Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters which originally load with the software when first initialized.

The present invention relates to microorganisms having increased plasmid copy number. Typically the plasmids will be those that are anti-sense RNA regulated including the following replicons: p15A and pMB1. Specifically, it has been discovered that mutations in five chromosomal genes, including *thrS, rpsA, rpoC, yjeR,* and *rhoL* resulted in the alteration of these classes of plasmids.

Plasmids

Plasmids are autonomous, self-replicating, extra-chromosomal elements generally not required for growth. Many of the genes on the plasmid allow for bacterial survival in a wide variety of challenging environments. Plasmids code for the proteins needed to initiate their replication. However, they do rely on the host cell replication machinery for replication. A "replicon" is any genetic element (e.g., plasmid, phage, cosmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*. A replicon comprises an origin of replication, to which another DNA segment may be attached so as to bring about the replication of the attached segment. Plasmids useful for gene expression are ubiquitous and well known in the art. Plasmids can be categorized based on several characteristics including copy number (single, low, medium, and high), method for regulation of copy number (iterons, AS-RNA, AS-RNA + repressor protein), method of replication (theta replication, strand displacement replication, rolling-circle replication) and incompatibility group. Plasmids derived from the same replicon replicate by the same mechanism and belong to the same incompatibility group.

Replication and control of circular bacterial plasmids is summarized in a review (del Solar et al., *Microbiol. And Mol. Boil. Rev.*, 62:434-464, (1998)). The first mechanism of plasmid copy number control is by iterons. The origin of replication for this class of plasmids, such as R6K, contain iterons. Iterons are directly repeated sequences necessary for replication and replication control. The iteron sites allow for the binding of the replication proteins that control plasmid replication. The second mechanism for copy control is by anti-sense RNA (AS-RNA). This is the mechanism by which ColE1 plasmids like p15A and pMB1 replicons are regulated. Briefly, inhibition of replication of these plasmids involves the interaction of RNA II, a post-transcriptionally processed transcript made by RNAP and RNA I, a 108-nucleotide anti-sense RNA complementary to the 5' end of RNA II. RNA I binds to RNA II and prevents its folding into a cloverleaf structure that is necessary for the formation of a stable RNA II/plasmid DNA hybrid for DNA synthesis. Rop is a small protein which when it binds to both RNA molecules, increases the stability of the RNA I/ RNA II complex, thus decreasing the likelihood of replication. The final method of copy control of plasmids, like R1, also involves AS-RNA. However, in these cases a transcriptional repressor, like CopG, interacts directly with the AS-RNA, RNA II. CopG binds to and represses

transcription of both the *copG* and *repB* genes. RNA II is a small RNA complementary to a region of the cop-rep mRNA. When the proteins have complexed with both the RNA and the AS-RNA then replication can not occur.

Target Genes

Plasmids can be used to express any endogenous or exogenous gene of interest for production of any desired genetic end product. Target genes may be drawn from a wide variety of biochemically important compounds including the pathways responsible for the synthesis of isoprenoids, carotenoids, terpenoids, tetrapyrroles, polyketides, vitamins, amino acids, fatty acids, proteins, nucleic acids, carbohydrates, antimicrobial agents, anticancer agents, poly-hydroxyalkanoic acid synthases, nitrilases, nitrile hydratases, amidases, enzymes used in the production of synthetic silk proteins, pyruvate decarboxylases, alcohol dehydrogenases, and biological metabolites.

For example suitable target genes will include, but are not limited to genes used in the production of poly-hydroxyalkanoic acid (PHA) synthases (*phaC*) which can be expressed for the production of biodegradable plastics, genes encoding nitrile hydratases for production of acrylamide, genes encoding synthetic silk protein genes for the production of silk proteins, the pyruvate decarboxylase gene (*pdc*), the alcohol dehydrogenase gene (*adh*) for alcohol production, genes encoding terpene synthases from plants for production of terpenes, genes encoding cholesterol oxidases for production of the enzyme,genes encoding monooxygenases derived from waste stream bacteria, the upstream isoprenoid pathways genes such as *dxs, dxr, ispA, ispD, ispE, ispF, lytB,* and *gcpE* to increase the flux of the isoprenoid pathway, the carotenoid synthesis and functionalization genes such as *crtE, crtB, crtI, crtY, crtW, crtO,* and *crtZ* to increase carotenoid production, genes used in tetrapyrrole biosynthesis, genes used in the production of polyketides, genes used in the synthesis of vitamins, genes used in the synthesis of fatty acids, genes used in the synthesis of carbohydrates, genes used in the production of antimicrobial agents, genes used in the synthesis of anti-canter agents, genes used in the synthesis of proteins and amino acids, genes used in the synthesis of nucleic acid, and genes used in the synthesis of biological metabolites. The preferred target genes used in the present invention are the *crtEXYIB* gene cluster from *Pantoea stewartii* ATCC 8199 (SEQ ID NOs. 1, 3, 5, 7, 9, and 11).

Optionally, one may produce the genetic end product as a secretion product of the transformed host. Secretion of desired proteins into the growth media has the advantages of simplified and less costly purification procedures. It is well known in the art that secretion signal sequences are often useful in facilitating the active transport of expressible proteins across cell membranes. The creation of a transformed host capable of secretion may be accomplished by the incorporation of a DNA sequence that codes for a secretion signal which is functional in the host production host. Methods for choosing appropriate signal sequences are well known in the art (EP 546049; WO 93/24631). The secretion signal DNA or facilitator may be located between the expression-controlling DNA and the instant gene or gene fragment, and in the same reading frame with the latter.

The plasmids or vectors may further comprise at least one promoter suitable for driving expression of genes in microbial hosts that will support the replication of the plasmids. Typically these promoters, including the initiation control regions, will be derived from native sources so that they function well in the preferred hosts. Termination control regions may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary, however, it is most preferred if included.

Carotenoid Biosynthesis

Carotenoids are pigments that are ubiquitous throughout nature and synthesized by all oxygen evolving photosynthetic organisms, and in some heterotrophic growing bacteria and fungi. Industrial uses of carotenoids include pharmaceuticals, food supplements, electro-optic applications, animal feed additives, and colorants in cosmetics, to mention a few. Because animals are unable to synthesize carotenoids *de novo*, they must obtain them by dietary means. Thus, manipulation of carotenoid production and composition in plants or bacteria can provide new or improved sources of carotenoids.

The genetics of carotenoid pigment biosynthesis are well known (Armstrong et al., *J. Bact.*, 176: 4795-4802 (1994); *Annu. Rev. Microbiol.* 51:629-659 (1997)). This pathway is extremely well studied in the Gram-negative, pigmented bacteria of the genera *Pantoea,* formerly known as *Erwinia*. In both *E. herbicola* EHO-10 (ATCC 39368) and *E. uredovora* 20D3 (ATCC 19321), the *crt* genes are clustered in two operons, *crt Z* and *crt EXYIB* (US 5,656,472; US 5,545,816; US 5,530,189; US 5,530,188; and US 5,429,939). Despite the similarity in operon structure, the DNA sequences of *E. uredovora* and *E. herbicola crt* genes show no homology by DNA-DNA hybridization

(US 5,429,939). The *Pantoea stewartii crt* genes have been described previously (US SN 10/218118; WO 02/079395).

Carotenoids come in many different forms and chemical structures. Most naturally occurring carotenoids are hydrophobic tetraterpenoids containing a C_{40} methyl-branched hydrocarbon backbone derived from successive condensation of eight C_5 isoprene units (isopentenyl diphosphate, IPP). In addition, novel carotenoids with longer or shorter backbones occur in some species of nonphotosynthetic bacteria.

E. coli contain the biosynthetic pathway necessary to synthesize farnesyl pyrophosphate (FPP) from IPP. FPP synthesis is common in both carotenogenic and non-carotenogenic bacteria. *E.coli* do not normally contain the genes necessary for conversion of FPP to β-carotene. Because of this, an *E.coli* strain containing a reporter plasmid (pPCB15) was used which has the additional genes necessary for β-carotene production in *E. coli* (Figure 1; SEQ ID NO: 28). Enzymes in the subsequent carotenoid pathway used to generate carotenoid pigments from FPP precursor can be divided into two categories: carotene backbone synthesis enzymes and subsequent modification enzymes. The backbone synthesis enzymes include geranyl geranyl pyrophosphate synthase (CrtE), phytoene synthase (CrtB), phytoene dehydrogenase (CrtI) and lycopene cyclase (CrtY/L), etc. The modification enzymes include ketolases, hydroxylases, dehydratases, glycosylases, etc.

Engineering *E. coli* for increased carotenoid production has previously focused on overexpression of key isoprenoid pathway genes from multi-copy plasmids. Various studies have report between a 1.5X and 50X increase in carotenoid formation in such *E. coli* systems upon cloning and transformation of plasmids encoding isopentenyl diphosphate isomerase (*idi*), geranylgeranyl pyrophosphate (GGPP) synthase (*gps*), deoxy-D-xylulose-5-phosphate (DXP) synthase (*dxs*), DXP reductoisomerase (*dxr*) from various sources (Kim, S.-W., and Keasling, J. D., *Biotech. Bioeng.*, 72:408-415 (2001); Mathews, P. D., and Wurtzel, E. T., *Appl. Microbiol. Biotechnol.*, 53:396-400 (2000); Harker, M, and Bramley, P. M., *FEBS Letter.*, 448:115-119 (1999); Misawa, N., and Shimada, H., *J. Biotechnol.,* 59:169-181 (1998); Liao et al., *Biotechnol. Bioeng.,* 62:235-241 (1999); and Misawa et al., *Biochem. J.*, 324:421-426 (1997)). In the present invention, the lower-carotenoid pathway genes *crtEXYIB*, rather the upper isoprenoid genes, were expressed on the multicopy plasmid. The chromosomal mutations described in the

present invention increase the copy number of the plasmids, thus increasing carotenoid production.

Mutations

Mutations isolated in this invention were all transposon insertions near the 3' end of essential genes, which likely resulted in altered gene function. These genes are involved in transcription and translation. Homologs of these genes are present in other organisms. Mutations of these homologs would be expected to have the same effect on these plasmids as mutations in *E. coli* genes.

The structural gene for threonyl-tRNA synthetase is *thrS*. It is one of the tRNA synthetases that bring together the specific amino acid it codes for and its tRNA molecule specific for that amino acid. It is an essential gene involved in protein synthesis. In the present invention, mutant Y1 contains a transposon disrupted *thrS* gene (SEQ ID NO. 19).

Ribosomal protein S1 is encoded by the *rpsA* gene. This protein facilitates the binding between the mRNA molecule and the ribosome. Ribosomes deficient in protein S1 are unable to extend the elongating peptide and are lethal to *E. coli*. However, one study demonstrated that a mutant lacking the 120 amino acids at the COOH-terminal region of the protein does not have significantly altered activity. Mutant Y8 contains a transposon disrupted *rpsA* gene (SEQ ID NO. 21).

The β' subunit of the RNA polymerase is encoded by the *rpoC* gene. It is an essential gene involved in transcription. Mutations near the 3' end of the gene were isolated and had a pleiotropic effect. A specific point mutation or a 3' end deletion of *rpoC* resulted in substantial reductions of the copy number of a ColE1 plasmid. (Ederth et al., *Mol Genet Genomics*, 267 (5): 587-592 (2002)). The *rpoC* 3' mutation by transposon insertion isolated in this invention had the opposite effect, increasing the copy number of the ColE1 plasmids. Mutant Y12 contains a transposon disrupted *rpoC* gene (SEQ ID NO. 23).

The gene *yjeR* (renamed *orn*) codes for an oligoribonuclease with a specificity for small oligoribonucleotides. Studies by Ghosh and Deutscher (*PNAS*, 96: 4372-4377 (1999)) indicate that the *yjeR* gene product is responsible for degrading small mRNA molecules to mononucleotides, a process necessary for cell viability. Mutant Y15 contains a transposon disrupted *yjeR* gene (SEQ ID NO. 24).

The leader peptide of the *rho* operon is encoded by the *rhoL* gene. The protein factor rho is responsible for terminating transcription at

specific sites of the RNA. In genes relying on this small protein for transcription termination, rho binds to the RNA causing the RNA polymerase to fall off of the DNA. Mutant Y17 contains a transposon disrupted *rhoL* gene (SEQ ID NO. 25).

Other mutations affecting plasmid copy number can be isolated using similar strategy as depicted in Figure 1. The reporter gene on the plasmid can be any gene that permits an easy visual screen. Examples of reporter genes include, but are not limited to *lacZ, gfp, lux, crt, xylMA*, etc. Selection strategy may also be designed such that only gene expression from certain range of copy number of plasmids will allow survival of the hosts.

Additionally, it can be envisioned that the reporter genes may be incorporated into plasmids containing different types of replicons. The present method could be used to identify chromosomal mutations that alter the plasmid copy number for each type of replicon tested.

Lastly, the identified disrupted genes may be used alone or in combination to genetically engineer bacteria for optimal plasmid expression useful for industrial production of a desired genetic end product.

Production Host

The ColE1-like plasmids can be used to produce any genetic end products in any hosts that will support their replication. Preferred production hosts include those that have the ability to harbor ColE1-like plasmids. The ColE1 plasmids have been reported to replicate in some other bacteria in addition to *Escherichia coli*. The pUC- and pBR-based cloning vectors (both ColE1 type plasmids) were shown to be maintained in *Pseudomonas stutzeri* (Pemberton et al., *Curr Microbiol*, 25:25-29 (1992)). Plasmids containing the p15A origin of replication can replicate freely in *Shewanella putrefaciens* (Myers et al., *Lett Appl Microbiol*, 24:221-225 (1997)). Plasmids very similar to ColE1 plasmids were also isolated from other bacteria such as *Salmonella enterica* (Astill et al., *Plasmid*, 30:258-267 (1993); *Erwinia stewartii* (Fu et al., *Plasmid*, 34:75-84 (1995); *Proteus vulgaris* (Koons et al., *Gene*, 157:73-79 (1995); and *Enterobacter agglomerans* (Mikiewicz et al., *Plasmid*, 38:210-219 (1997)). Additional bacteria capable of supporting ColE1-like plasmids include *Actinobacillus* sp., *Yersinia* sp., and *Pantoea* sp. Most preferred production hosts are enteric production hosts, particularly those of the genera *Escherichia* and *Salmonella*.

Enteric bacteria are members of the family *Enterobacteriaceae* and include such members as *Escherichia*, *Salmonella*, and *Shigella*. They are gram-negative straight rods, 0.3-1.0 X 1.0-6.0 mm, motile by peritrichous flagella (except for *Tatumella*) or nonmotile. They grow in the presence and absence of oxygen and grow well on peptone, meat extract, and (usually) MacConkey's media. Some grow on D-glucose as the sole source of carbon, whereas others require vitamins and/or mineral(s). They are chemoorganotrophic with respiratory and fermentative metabolism but are not halophilic. Acid and often visible gas is produced during fermentation of D-glucose, other carbohydrates, and polyhydroxyl alcohols. They are oxidase negative and, with the exception of *Shigella dysenteriae* 0 group 1 and *Xenorhabdus nematophilus*, catalase positive. Nitrate is reduced to nitrite (except by some strains of *Erwinia* and *Yersina*). The G + C content of DNA is 38-60 mol% (T_m, Bd). DNAs from species within most genera are at least 20% related to one another and to *Escherichia coli*, the type species of the family. Notable exceptions are species of *Yersina, Proteus, Providenica, Hafnia* and *Edwardsiella*, whose DNAs are 10-20% related to those of species from other genera. Except for *Erwinia chrysanthemi*, all species tested contain the enterobacterial common antigen (Bergy's Manual of Systematic Bacteriology, D. H. Bergy et al., Baltimore: Williams and Wilkins, 1984).

General methods for introducing plasmids into these preferred hosts include chemical-induced transformation, electroporation, conjugation and transduction. The preferred hosts can be grown in tryptone yeast extract based rich media, or defined media with all the essential nutrients. Suitable antibiotics can be added in the growth media to maintain the plasmids. Similar gene mutations in the preferred hosts are expected to have similar effect of increasing copy number of the ColE1 and like plasmids replicated in these hosts.

DESCRIPTION OF PREFERRED EMBODIMENTS

Five mutant genes have been identified in *E. coli* which unexpectedly had effects on plasmid copy number. In particular, transposon mutagenesis of genes *thrS*, *rpsA*, *rpoC*, *yjeR*, and *rhoL* resulted in an increase of plasmid copy number of certain plasmids. The plasmids effected were those exhibiting anti-sense RNA copy-number control including those using the pMB1 and p15A replicons.

In one embodiment, the *crt* carotenoid biosynthesis gene cluster from *Pantoea stewartii* (ATCC No. 8199) was cloned, sequenced, and

characterized (Examples 1 and 2; Tables 1 and 2). A reporter plasmid (pPCB15; SEQ ID NO. 28) was created which functionally expressed the *crtEXYIB* gene cluster (Example 3). The reporter plasmid was transformed into *E. coli* MG1655, enabling the strain to produce β-carotene (yellow colonies).

In another embodiment, transposon mutagenesis was conducted on *E. coli* MG1655 (pPCB15) (Figure 1). Mutant colonies appearing to have a phenotypic color change (either deeper yellow or white appearance) were isolated and characterized. The level of β-carotene production was measured spectrophotometrically and verified by HPLC analysis (Example 3). The pigment yield was measured relative to the control strain harboring only the pPCB15 reporter plasmid (Figure 2). Mutants Y4, Y15, Y16, Y17, and Y21 exhibited a 1.5-2 fold increase in β-carotene productionMutants Y1, Y8, and Y12 exhibited a 2.5-3.5 fold increase in β-carotene production. The chromosomal transposon insertion sites in the *E. coli* mutants were identified and sequenced (Example 4; Table 3).

In another embodiment, the increased carotenoid production in the mutant strains was attributed to an increase in reporter plasmid copy number. The reporter plasmid copy number was measured in the mutants (Example 5; Figures 3 and 4). Mutants Y1, Y8, Y12, Y15, and Y17 have a 2-4 fold increase in plasmid DNA when compared to the control Mutants Y4, Y16, and Y21 had comparable amounts of plasmid DNA to the control while mutant W4 had much less plasmid DNA (Figures 3 and 4).

In another embodiment, the increased in plasmid copy number was generally attributed to plasmids having ColE1-type replicons and was not specifically associated with the pPCB15 reporter plasmid. The pPCB15 reporter plasmid was cured from the mutants and different pACYC-derived plasmids were tested (Example 6). Plasmid pTV200, containing a *luxCDABE* reporter construct, was transformed into the various cured mutant strains. The lux activity was decreased 60% in W4, whereas it increased 4 to 7 fold in Y1 and Y8 mutants and over 10 fold in Y12, Y15, and Y17 mutants (Figure 5). Plasmid pTV200 copy number was determined and was consistent with the change of luciferase activity (Figure 6).

In another embodiment, the various mutant strains were shown to affect plasmids harboring p15A and pMB1 replicons. Plasmid pBR328 (pMB1 replicon) was transformed into cured mutant hosts Y1, Y8, Y12,

Y15, and Y17. Plasmid pBR328 DNA levels were analyzed from the mutant hosts and were found to be increased approximately 2-4 fold above control levels (Example 7; Figure 7). Various other plasmids with different replicon types were analyzed in mutant hosts W4 and Y15 versus the control (Example 8; Table 4). The increased copy number associated with the various mutations was not observed in plasmids harboring replicons other than pMB1 and p15A. The mutations observed in Y1, Y8, Y12, Y15 and Y17 were shown to increase plasmid copy number in plasmids with p15A and pMB1 replicons.

In another embodiment, reporter plasmids containing different replicons can be created and used to identify chromosomal mutations that increase plasmid copy number. The *Pantoea stewartii crtEXYIB* gene cluster could be cloned and expressed in reporter plasmids containing different replicons. Transposon mutagenesis could be used to identify mutations associated with each replicon type.

In another embodiment, the present method could be used to identify additional genes associated with increasing plasmid copy number in those plasmids having p15A and pMB1 replicons. These mutations, as well as those identified in the present invention, could be used alone or in combination to genetically engineer increased production of a desired genetic end product. In a preferred embodiment, the mutation information could be used to engineer *E. coli* strains for increased production of carotenoids.

EXAMPLES

The present invention is further defined in the following Examples. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

GENERAL METHODS

Standard recombinant DNA and molecular cloning techniques used in the Examples are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, (1989) (Maniatis) and by T. J. Silhavy, M. L. Bennan, and L. W. Enquist,

Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and by Ausubel, F. M. et al., Current Protocols in Molecular Biology, pub. by Greene Publishing Assoc. and Wiley-Interscience (1987).

Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in Manual of Methods for General Bacteriology (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds), American Society for Microbiology, Washington, DC. (1994)) or by Thomas D. Brock in Biotechnology: A Textbook of Industrial Microbiology, Second Edition, Sinauer Associates, Inc., Sunderland, MA (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of bacterial cells were obtained from Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), or Sigma Chemical Company (St. Louis, MO) unless otherwise specified.

Manipulations of genetic sequences were accomplished using the suite of programs available from the Genetics Computer Group Inc. (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI). Where the GCG program "Pileup" was used the gap creation default value of 12, and the gap extension default value of 4 were used. Where the CGC "Gap" or "Bestfit" programs were used the default gap creation penalty of 50 and the default gap extension penalty of 3 were used. Multiple alignments were created using the FASTA program incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.,* [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Publisher: Plenum, New York, NY). In any case where program parameters were not prompted for, in these or any other programs, default values were used.

The meaning of abbreviations is as follows: "h" means hour(s), "min" means minute(s), "sec" means second(s), "d" means day(s), "μL" means microliter(s), "mL" means milliliter(s), "L" means liter(s), and "rpm" means revolutions per minute.

EXAMPLE 1

Cloning of β-Carotene Production Genes from *Pantoea stewartii*

Primers were designed using the sequence from *Erwinia uredovora* to amplify a fragment by PCR containing the *crt* genes. These sequences included 5'-3':

ATGACGGTCTGCGCAAAAAAACACG	SEQ ID 13
GAGAAATTATGTTGTGGATTTGGAATGC	SEQ ID 14

Chromosomal DNA was purified from *Pantoea stewartii* (ATCC NO. 8199) and *Pfu* Turbo polymerase (Stratagene, La Jolla, CA) was used in a PCR amplification reaction under the following conditions: 94°C, 5 min; 94°C (1 min)-60°C (1 min)-72°C (10 min) for 25 cycles, and 72°C for 10 min. A single product of approximately 6.5 kb was observed following gel electrophoresis. *Taq* polymerase (Perkin Elmer, Foster City, CA) was used in a ten minute 72°C reaction to add additional 3' adenosine nucleotides to the fragment for TOPO cloning into pCR4-TOPO (Invitrogen, Carlsbad, CA) to create the plasmid pPCB13. Following transformation to *E. coli* DH5α (Life Technologies, Rockville, MD) by electroporation, several colonies appeared to be bright yellow in color indicating that they were producing a carotenoid compound. Following plasmid isolation as instructed by the manufacturer using the Qiagen (Valencia, CA) miniprep kit, the plasmid containing the 6.5 kb amplified fragment was transposed with pGPS1.1 using the GPS-1 Genome Priming System kit (New England Biolabs, Inc., Beverly, MA). A number of these transposed plasmids were sequenced from each end of the transposon. Sequence was generated on an ABI Automatic sequencer using dye terminator technology (US 5,366,860; EP 272007) using transposon specific primers. Sequence assembly was performed with the Sequencher program (Gene Codes Corp., Ann Arbor, MI).

EXAMPLE 2

Identification and Characterization of *Pantoea stewartii* Genes

Genes encoding *crtE, X, Y, I, B,* and *Z*, cloned from *Pantoea stewartii,* were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul et al., *J. Mol. Biol.*, 215:403-410 (1993)) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank® CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The sequences obtained were analyzed for similarity to all

publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D., *Nature Genetics*, 3:266-272 (1993)) provided by the NCBI.

All comparisons were done using either the BLASTNnr or BLASTXnr algorithm. The results of the BLAST comparison is given in Table 2 which summarize the sequences to which they have the most similarity. Table 2 displays data based on the BLASTXnr algorithm with values reported in expect values. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

TABLE 2

ORF Name	Gene Name	Similarity Identified	SEQ ID base	SEQ ID Peptide	% Identity[a]	% Similarity[b]	E-value[c]	Citation
1	*crtE*	Geranylgeranyl pryophosphate synthetase (or GGPP synth or farnesyltranstransferase) EC 2.5.1.29 gi|117509|sp|P21684|CRTE_PANAN GERANYLGERANYL PYROPHOSPHATE SYNTHETASE (GGPP SYNTHETASE) (FARNESYLTRANSTRANSFERASE)	1	2	83	88	e-137	Misawa et al., *J. Bacteriol.* 172 (12), 6704-6712 (1990)
2	*crtX*	Zeaxanthin glucosyl transferase EC 2.4.1.- gi|1073294|pir||S52583 crtX protein - *Erwinia herbicola*	3	4	75	79	0.0	Lin et al., Mol. *Gen. Genet.* 245 (4), 417-423 (1994)
3	*crtY*	Lycopene cyclase gi|1073295|pir||S52585 lycopene cyclase - *Erwinia herbicola*	5	6	83	91	0.0	Lin et al., *Mol. Gen. Genet.* 245 (4), 417-423 (1994)
4	*crtI*	Phytoene desaturaseEC 1.3.-.- gi|1073299|pir||S52586 phytoene dehydrogenase (EC 1.3.-.-) - *Erwinia herbicola*	7	8	89	91	0.0	Lin et al., *Mol. Gen. Genet.* 245 (4), 417-423 (1994)
5	*crtB*	Phytoene synthaseEC2.5.1.- gi|1073300|pir||S52587 prephytoene pyrophosphate synthase - *Erwinia herbicola*	9	10	88	92	e-150	Lin et al., *Mol. Gen. Genet.* 245 (4), 417-423 (1994)

ORF Name	Gene Name	Similarity Identified	SEQ ID base	SEQ ID Peptide	% Identity[a]	% Similarity[b]	E-value[c]	Citation
6	*crtZ*	Beta-carotene hydroxylase gi|117526|sp|P21688|CRTZ_PANAN BETA-CAROTENE HYDROXYLASE	11	12	88	91	3e-88	Misawa et al., *J. Bacteriol.* 172 (12), 6704-6712 (1990)

[a]%Identity is defined as percentage of amino acids that are identical between the two proteins.
[b]% Similarity is defined as percentage of amino acids that are identical or conserved between the two proteins.
[c]Expect value. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

EXAMPLE 3

Isolation of Chromosomal Mutations that Affect Carotenoid Production

Wild type *E. coli* is non-carotenogenic and synthesizes only the farnesyl pyrophosphate precursor for carotenoids. When the *crtEXYIB* gene cluster from *Pantoea stewartii* was introduced into *E.coli*, β-carotene was synthesized and the cells became yellow. *E. coli* chromosomal mutations which increase carotenoid production should result in deeper yellow colonies. *E. coli* chromosomal mutations which decrease carotenoid production should result in lighter yellow or white colonies (Figure 1).

The β-carotene reporter plasmid, pPCB15 (cam^R), encodes the carotenoid biosynthesis gene cluster (*crtEXYIB*) from *Pantoea Stewartii* (ATCC NO. 8199). The pPCB15 plasmid (SEQ ID NO. 28) was constructed from ligation of *Sma*I digested pSU18 (Bartolome et al., *Gene*, 102:75-78 (1991)) vector with a blunt-ended *Pme*I/*Not*I fragment carrying *crtEXYIB* from pPCB13 (Example 1). *E. coli* MG1655 transformed with pPCB15 was used for transposon mutagenesis. Mutagenesis was performed using EZ:TN™ <KAN-2>Tnp Transposome™ kit (Epicentre Technologies, Madison, WI) according to manufacturer's instructions. A 1 μL volume of the transposome was electroporated into 50 μL of highly electro-competent MG1655(pPCB15) cells. The mutant cells were spread on LB-Noble Agar (Difco laboratories, Detroit, MI) plates with 25 μg/mL kanamycin and 25 μg/mL chloramphenicol, and grown at 37°C overnight. Tens of thousands of mutant colonies were visually examined for deeper or lighter color development. The candidate mutants were re-streaked and frozen for further characterization.

To confirm if the deeper or lighter color colonies were indeed indicative of amount of β-carotene production, the carotenoids in the candidate mutants were extracted and quantified spectrophotometrically. Each candidate clone was cultured in 10 mL LB medium with 25 μg/mL chloramphenicol in 50 mL flasks overnight shaking at 250 rpm. MG1655(pPCB15) was used as the control. Carotenoid was extracted from each cell pellet for 15 min into 1 mL acetone, and the amount of β-carotene produced was measured at 455 nm. Cell density was measured at 600 nm. OD455/OD600 was used to normalize β-carotene production for different cultures. β-carotene production was also verified by HPLC. The averages of three independent measurements with standard deviations are shown in Figure 2. Among the mutant clones tested, eight

showed increased β-carotene production. Mutants Y1, Y8 and Y12 showed 2.5-3.5 fold higher β-carotene production. Mutants Y4, Y15, Y16, Y17 and Y21 showed 1.5-2 fold higher β-carotene production. Mutant W4 was a white mutant that decreased β-carotene production to 17% of that of the MG1655(pPCB15) control.

EXAMPLE 4

Mapping of the Transposon Insertions in *E. coli* Chromosome

The transposon insertion site in each mutant was identified by PCR and sequencing directly from the chromosome. A modified single-primer PCR method (Karlyshev et al., *BioTechniques*, 28:1078-82 (2000)) was used. A 100 μL volume of culture grown overnight was heated at 99°C for 10 min in a PCR machine. Cell debris was removed at 4000 g for 10 min. A 1 μL volume of the supernatant was used in a 50 μL PCR reaction using either Tn5PCRF (5'-GCTGAGTTGAAGGATCAGATC-3';SEQ ID 15) or Tn5PCRR (5'-CGAGCAAGACGTTTCCCGTTG-3';SEQ ID 16) primer. PCR was carried out as follows: 5 min at 95°C; 20 cycles of 92°C for 30 sec, 60°C for 30 sec, 72°C for 3 min; 30 cycles of 92°C for 30 sec, 40°C for 30 sec, 72°C for 2 min; 30 cycles of 92°C for 30 sec, 60°C for 30 sec, and 72°C for 2 min. A 10 μL volume of each PCR product was checked on an agarose gel. A 40 μL volume of each PCR product was purified using Qiagen PCR cleanup kit, and sequenced using sequencing primers Kan-2 FP-1 (5'-ACCTACAACAAAGCTCTCATCAACC-3';SEQ ID 17) or Kan-2 RP-1 (5'-GCAATGTAACATCAGAGATTTTGAG-3';SEQ ID 18) provided by the EZ:TN™ <KAN-2>Tnp Transposome™ kit. The chromosomal insertion site of the transposon was identified as the junction between the Tn5 transposon and MG1655 chromosome DNA by aligning the sequence obtained from each mutant with the *E. coli* genomic sequence. Table 3 summarizes the chromosomal insertion sites of the mutants. The numbers refer to the standard base pair (bp) numbers for *E. coli* genome of MG1655 (GenBank® Accession No. U00096). Majority of the genes affected are involved in transcription, translation or RNA stability. Five of them (*thrS, rpsA, rpoC, yjeR, rhoL*) were previously reported to be essential. The transposon insertions we obtained in these genes were very close to the carboxyl terminal end and most likely resulted in functional although truncated proteins.

Table 3. Localization of the transposon insertions in *E. coli* chromosome

Mutant	Transposon Insertion Site	Gene disrupted Location on *E. coli* chromosome	Gene Function	Essentiality reported	Reference
W4	158904	*pcnB*: 157729-159093	poly(A) polymerase	No	Masters M, 1993 *J Bacteriol*175:
Y1	1798679	*thrS*: 1798666-1800594	threonyl-tRNA synthetase	Yes	Johnson EJ, 1977 *J Bacteriol* 129:66-70
Y4	3304788	*deaD*: 3303612-3305552	RNA helicase	No	Toone WM, 1991 *J Bacteriol* 173:3291-302
Y8	962815	*rpsA*: 961218-962891	30S ribosomal subunit protein S1	Yes	Kitakawa M, 1982 *Mol Gen Genet* 185:445-7
Y12	4187062	*rpoC*: 4182928-4187151	RNA polymerase β' subunit	Yes	Post,L.E, 1979 *PNAS* 76:1697-1701
Y15	4389704	*yjeR*: 4389113-4389727	oligo-ribonuclease	Yes	Ghosh S, 1999 *PNAS* 96:4372-7.
Y16	3396592	*mreC*: 3396512-3397615	rod shape-determining protein	No	Wachi M, 1987 *J Bacteriol* 169:4935-40
Y17	3963892	*rhoL*: 3963846-3963947	rho operon leader peptide	Yes	Das A, 1976 *PNAS* 73:1959-63
Y21	2657233	*yfhE* (*hscB*): 2656972-2657487	heat shock cognate protein	Unknown	Takahashi Y, 1999 *J Biochem (Tokyo)*126: 917-26

EXAMPLE 5

Analysis of Plasmid Copy Number in the Mutants Affecting β-carotene Production

White mutant W4 had a transposon insertion in *pcnB* gene (SEQ ID NO. 27), which encodes a poly(A) polymerase that polyadenylates RNA. Mutation in *pcnB* gene was reported to decrease the copy number of ColE1 plasmids. It was possible that the effect on carotenoid production in some of the mutants was due to copy number change of the carotenoid-synthesizing plasmid. We analyzed the amount of the β-carotene synthesizing plasmid pPCB15, a derivative of pACYC plasmids, in the isolated mutants. Cells were grown in LB containing chloramphenicol (25 μg/mL) with shaking overnight. Cell density was measured by OD600. Plasmid DNA was isolated from same amount of cells (not the same volume) using Qiagen miniprep spin kit. A 5-μL volume of *Eco*RI-digested plasmid DNA isolated from each strain was loaded on an agarose gel for comparison. Figure 3 shows the plasmid DNA isolated from two independent clones of each stain. In both experiments, Mutants Y1, Y8, Y12, Y15 and Y17 appeared to have more plasmid DNA than wild type MG1655. Mutant W4 had much less plasmid DNA. Mutants Y4, Y16 and Y21 had comparable amount of plasmid DNA as MG1655. To estimate the change of the plasmid copy number 1 μL, 2 μL and 4 μL of digested DNA from Y1, Y8, Y12, Y15, Y17 and MG1655 were loaded on an agarose gel as shown in Figure 4. All five mutants showed a 2 to 4 fold increase in plasmid copy number compared to MG1655. It is interesting to note that these five mutants all contained mutations in an essential gene. A recent report described a different mutation of *rpoC* from that in Y12 mutant that decreased the copy number of ColE1 plasmids (Ederth et al., *Mol. Gen. Genomics*, 267:587-592 (2002)).

EXAMPLE 6

Luciferase Expression in *E. coli* Mutants that Affect Plasmid Copy Number

To determine if the copy number effect was specifically associated with the carotenoid-synthesizing plasmid or not, the pPCB15 (Cam^R) plasmid was cured from the mutants. A different pACYC-derived plasmid was tested in the cured strains. The plasmid-cured strains were isolated by growing the cells in the absence of chloramphenicol and plating dilutions on LB plates containing kanamycin. The kanamycin resistant colonies that became chloramphenicol sensitive had presumably lost the pPCB15 plasmid.

Plasmid pTV200 contains a promoterless *luxCDABE* from *P. luminescens* in pACYC184. *E. coli* strains containing pTV200 are positive

for luciferase (lux) activity, presumably due to expression from the chloramphenicol resistance gene promoter on pACYC184 vector. We tested luciferase expression of pTV200 in different cured strains.

Bacterial bioluminescence is a phenomenon in which the products of 5 structural genes (*luxA*, *luxB*, *luxC*, *luxD,* and *luxE*) work in concert to produce light. The *luxD* product generates a C14 fatty acid from a precursor. The C14 fatty acid is activated in an ATP dependent reaction to an acyl-enzyme conjugate through the action of the *luxE* product, which couples bacterial bioluminescence to the cellular energetic state. The acyl-enzyme (*luxE* product) serves as a transfer agent, donating the acyl group to the *luxC* product. The acyl-LuxC binary complex is then reduced in a reaction in which NADPH serves as an electron pair and proton donor reducing the acyl conjugate to the C14 aldehyde. This reaction couples the reducing power of the cell to bacterial light emission. The light production reaction, catalyzed by luciferase (the product of *luxA* and *luxB*), generates light. The energy for light emission is provided by the aldehyde to fatty acid conversion and $FMNH_2$ oxidation, providing another couple between light production and the cellular energy state.

The *Photorabdus luminenscens luxAB* genes were used as reporters for plasmid copy number alterations via the mutated genes (Van Dyk et al., *Appl. Environ. Microbiol.*, 180:785-792 (1995)). Plasmid pTV200 is a pACYC184-derived plasmid carrying the *Photorhabdus luminescens luxCDABE* operon. It was constructed in the following manner. Plasmid pJT205 (Van Dyk, T., and Rosson, R., *Photorhabulus luminescens luxCDABE* promoter probe vectors, in Method in Molecular Biology: Bioluminescence Methods and Protocols, Vol. 102, LaRossa, R.A., Ed., Humana Press Inc., Towowa, NJ, pp. 85 (1998)) was digested with restriction enzymes *Eco*RI and *Pvu*II. The products of this digestion were ligated with plasmid pACYC184 that had been digested with the same two enzymes. The ligation mixture was used to transform *E. coli* strain DH5α, selecting for tetracycline resistance. The agar plates containing the transformant colonies were use to expose Kodak XAR film and colonies that produced light were purified. The light-producing colonies were screened for sensitivity to ampicillin and chloramphenicol. Plasmid DNA was obtained from one tetracycline-resistant, light-producing, ampicillin-sensitive, and chloramphenicol-sensitive isolate. This plasmid, named pTV200, had two bands of the expected size following *Bam*HI digestion.

Plasmid pTV200 was transformed into the plasmid-cured mutant strains with tetracycline selection. Luciferase activity and pTV200 plasmid concentration were analyzed from the mutants. Cells containing pTV200 were grown in LB with 10 μg/mL of tetracycline at 37°C with shaking overnight. A 100-μL volume of each cell culture was pipetted into a 96-well plate and luciferase activity was measured using HTS 7000 Plus BioAssay Reader (Perkin Elmer, Norwalk, CT). Cell density of the samples in each well was also measured by absorption at OD600 using the BioAssay Reader. The normalized luciferase activity of each sample was calculated and is shown in Figure 5. The lux activity decreased 60% in W4 mutant compared to the wild type MG1655, whereas it increased 4 to 7 fold in Y1 and Y8 mutants and over 10 fold in Y12, Y15 and Y17 mutants. Plasmid DNA was also isolated from same amount of cells for different strains and digested with *Eco*RI. Aliquots (2 μL and 4 μL) of digested plasmid DNA were loaded on agarose gels. Comparison of pTV200 isolated from wild type MG1655 and mutants is shown in Figure 6. Consistent with the luciferase activity assay, the copy number of pTV200 decreased in the W4 mutant, whereas it increased in Y1, Y8, Y12, Y15 and Y17 mutants.

EXAMPLE 7

Effect of the Chromosomal Mutations on the Copy Number of Plasmids of pMB1 Replicon

It is known that *pcnB* mutation affects copy number of plasmids of both p15A and pMB1 replicons (Liu et al., *supra*). We tested if the other mutations we isolated also affected copy number of pMB1-derived plasmids. Plasmid pBR328 (pMB1 replicon) was transformed into cured mutant hosts of Y1, Y8, Y12, Y15 and Y17. Plasmid DNA was isolated from same amount of cells from each strain and digested with *Eco*RI. Aliquots (2 μL and 4 μL) of digested plasmid DNA were loaded on agarose gels. As shown in Figure 7, the copy number of pBR328 increased approximately 2-4 fold in Y1, Y8, Y12, Y15 and Y17 mutants comparing to that in the wild type MG1655.

EXAMPLE 8

Effect of the Chromosomal Mutations on the Copy Number of Plasmids of Different Replicons

To determine if the above mutations would affect the copy number of other plasmids, we tested a list of plasmids of different replicons (Table 4) in these mutant hosts. A representative of the mutants that increase the plasmid

copy number, Y15, and the W4 mutant that decreased the plasmid copy number were used for this experiment. Plasmids shown in Table 4 were transformed into MG1655 and the cured mutant hosts of W4 and Y15, and selected with the respective antibiotics. The cells were grown in LB containing the appropriate antibiotics and plasmid DNA was prepared from the same amount of cells using Qiagen miniprep spin columns (Qiagen, Inc., Carlsbad, CA). Plasmid DNA was digested with *Eco*RI and aliquots of the digested DNA (1 μL, 2 μL, 4 μL, and 16 μL) were loaded on an agarose gel (Figure 8). The *pcn*B (in W4) or *yje*R (in Y15) mutation did not appear to affect the copy number of pSC101, pBHR1, pMMB66 and pTJS75. The *pcn*B mutation decreased the copy number of pBR328 and pACYC184 more than 16 fold. The *yje*R mutation increased the copy number of pBR328 and pACYC184 about 2 fold. Therefore, these *E. coli* chromosomal mutations affected the copy number of plasmids with replicons pMB1 and/or p15A.

Table 4. Plasmids of different replicons tested in the W4 and Y15 mutant hosts

Plasmid	Replicon	Antibiotic marker	Reference
pBR328	pMB1	Cm^r Ap^r Tc^r	Balbas, P. 1986 *Gene* 50:3-40
pACYC184	p15A	Cm^r Tc^r	Chang, ACY. 1978 *J. Bacteriol.* 134:1141-56
pSC101	pSC101	Tc^r	Cohen, SN. 1977 *J. Bacteriol.* 132:734-737
pBHR1	pBBR1	Cm^r Kn^r	Antoine, R. 1992 *Mol Microbiol.* 6:1785-99
pMMB66	RSF1010 (IncQ)	Ap^r	Scholz, P. 1989 *Gene* 75:271-288
pTJS75	RK2 (IncP)	Tc^r	Schmidhauser, TJ. 1985 *J. Bacteriol.* 164:446-455